Guidance on Research Data Handling for PGR Students and Staff*

This document is intended to assist in completing the research data management section of the SSP Application for Research Ethics Approval for **PGR students and Staff, and by taught students where their work may lead to an academic publication**. There is a separate version of this document specifically for use by taught students only, where their work will not lead to a publication.

Note that this document focusses on the typical research projects in SSP. Some other types of research (e.g. involving NHS staff, patients or patient data, or animals) may have their own specific guidelines.

General Requirements

The University of Nottingham Research Data Management Policy, https://www.nottingham.ac.uk/fabs/rgs/research-data-management/creating-data/policies.aspx

See also funder requirements where relevant.

Will **personal data** (including photos, video or audio) be collected, recorded or used?

In general, personal data is data from or about a living individual, who might be identifiable from that data (perhaps when combined with other available data). There is further guidance on identifying personal data on the SSP-REC website.

> This may include data that you capture, such as video or audio recordings.
>
> It may also include data that you get from elsewhere, from which an individual might be identified. For example, you still need to consider the ethics of using data obtained from social media sites or other researchers.

What data (or materials) will be collected or used?

Identify the type of data and also the subject or content of the data.

> E.g. video of participants testing a user interface, questionnaire responses, logs of network traffic from a home network, field notes of direct observations of…, audio recordings of interviews.

What if any **constraints** apply to use of this data (or materials)?

In general participants should give (or have already given) informed consent to use of their data that includes what you propose to do.

> If the data was collected by someone else, then check what consent was given when it was collected.
>
> If the data is obtained from a third party such as a social media site, then it will typically also be subject to a license agreement that restricts what can be done with it.

* Adapted from guidance by Chris Greenhalgh, School of Computer Science

Some data may also be subject to copyright, e.g. the words that people say or write. You need to ensure that your use also respects this and/or permission has been obtained.

Some data may also be company confidential, e.g. internal details of a business process or commercial operation, obtained under a contract (e.g. a non-disclosure agreement). Again, note these constraints if they apply (typically your supervisor will be able to advise in this case).

How will this data (or materials) be **collected or obtained**?

Explain briefly how the data will be obtained. For example, will you collect or record it yourself? Where/when?

Consider whether this raises any additional complications, e.g. might you also record other people who are not part of your research and who haven't given consent?

Note that for online surveys you still need to comply with GDPR/data protection requirements.

> For simple surveys you should be able to use the University-provided Office365 'forms' functionality.
>
> For more complex surveys, staff and PGR students can use the free Jisc Online Surveys, administered by the graduate school – see https://www.nottingham.ac.uk/graduateschool/traininganddevelopment/resources/bristolonlinesurveys.aspx
>
> Try to make survey responses anonymous if possible and include information about the study at the start of the survey with a compulsory checkbox to say that they understand and agree to participate.

How will this data (or materials) be **processed before analysis**?

What if any processing will be done on the data before it is analysed?

> For example, audio recordings may be transcribed.

Also make clear whether/what anonymization takes place at this stage.

> For example, use pseudonyms or IDs for participants, not their real names; removing identifying information from free text responses.

How will this data (or materials) be **stored and secured** (a) **during the research** (b) **after the research**?

In general, you should use Teams to store and share research data, including personal data, with other members of the University (quota 1TB – 25TB/team).

> Alternatively you can use the University-provided Office 365 OneDrive to store research data, including personal data (up to 5TB/user). However, be aware that storage in OneDrive is probably linked to the individual who created it (even if it is shared), and if they leave the

University then it will be deleted. So in that case copies may need to be maintained by others, e.g. supervisor, other research group members.

See also the IS page on managing research data, https://www.nottingham.ac.uk/it-services/research/data/index.aspx

The University security policy and Restricted Data Handling Policy[†] classify data as:

University Restricted Data, i.e. personal data and commercially sensitive data, which needs security and fine-grained access control.

University Standard Data, is freely available to all UoN staff and students

Public data, is freely accessible to all parties.

For restricted data (i.e. personal data and commercially sensitive data):

All restricted data must be recorded (by the School) in a data asset inventory, including title, restricted data fields, owner, stewards (and responsibilities), users and where the asset is stored. This will be done via the Ethics application.

Restricted data must be secured appropriately, e.g.

User access must be recorded, controlled and reviewed periodically. Devices holding data must be appropriately protected (e.g. by password and/or physical security). Access must require identification AND authentication. Data transfer must be encrypted. Third party / cloud mechanisms can only be used where specifically approved against the policy. Portable storage must have a secure physical backup. Devices/media without encryption must not be taken outside the University. Laptops and mobile devices must be protected at all times, e.g. encrypted.

Where personal data cannot be encrypted (e.g. as it is being captured on a video camera or voice recorder) you should plan to transfer it to a securer form of storage as soon as possible.

Any sensitive personal data (e.g. health, sexuality) MUST be encrypted (just being stored on a 'secure' store is not enough).

How will the data be **formatted**, i.e. what data formats or encodings (e.g. file types, media encodings) will be used?

In general, consider using open formats that are likely to still be usable/readable in 5-10 years' time.

You may have working copies in other (e.g. tool-specific) proprietary formats while you are actively working with the data but consider also keeping copies in standard or open formats, even if these lose some details.

[†] See UoN page on information security, https://www.nottingham.ac.uk/it-services/security/strategy/it-security.aspx including the (2015/16) information security policy and the (more recent) University Restricted Data Policy, 'University Data Handling Standards Policy' from https://workspace.nottingham.ac.uk/display/GDPR/Data+Protection

How will the data be **organised**?

E.g. what folder structure will you use, what metadata will you provide?

It may be useful to subdivide the data into:

1. **Method and approvals** (University or Public) – e.g. ethics forms, information sheet, etc. These will NOT usually be restricted data.
2. **Participants** (Restricted) – i.e. completed (scanned) consent forms, participant lists, records of withdrawals (and steps taken). These will be restricted data.
3. **Raw Data** (Restricted) – i.e. sensitive or personal data such as audio/video recordings, logs with location or physiological data, as captured. There may be several sub-folders, e.g. for data type or session. Details of anonymization should be recorded here (if relevant). Once data has been collected this folder should not normally be changed.
4. **Working Data** (Restricted, University or Public) – i.e. data as used for the main research activity, anonymised where possible. There may be several sub-folders, e.g. for data type or session. Once data has been collected this folder should not normally be changed (although synthetic data that can be regenerated may be deleted).
5. **Analysis and Results** (Restricted, University or Public) – i.e. work being done on/with the data. There may be several sub-folders, e.g. for particular researchers, analyses or reports. This may continue to change during the active research phase.
6. **Archive Data** (Restricted, University or Public) – i.e. data to be retained specifically for future use. There may be several sub-folders if different parts or aspects of the data are archived separately or at different times.
7. **Published Data** (Public) – i.e. data that has been published, e.g. in or underlying a publication. There may be several sub-folders if different parts or aspects of the data are published separately or at different times.

(Omit sections that are not relevant)

Making the security classification explicit will probably help in the future (i.e. Restricted, University [Standard] or Public). The separation also helps with managing retention and deletion, e.g. 1-5 will be retained for 7 years, while 6 and 7 may retained for longer (see below); and 7 may be copied to a public repository for longer-term archiving (see below).

Where there are different parts to the research (e.g. related studies) it is probably better to reflect this at the top level and use the above sub-divisions within each sub-study. Where data from multiple studies is combined, identify this as a separate top-level activity.

Typically, each data folder should contain basic metadata, i.e. a short description of the data that it contains, source, subject and format(s). For example, this might be in a textual 'README' file in each directory.

How will this data (or materials) be **analysed**?

What kinds of methods will you use to analyse the data? What tools will you use?

You don't need to go into lots of detail but try to make clear whether the results will potentially include personal data or not. For example, statistics from a population will generally not be personal data, but fragments of video featuring recognisable individuals will be.

How will this data (or materials) be **reported** in publications?

You don't need to give a lot of detail but make it clear what kinds of results might be published and how personal these might be, as above.

In most cases publications will be anonymous.

Where users are referred to specifically, pseudonyms or study-specific identifiers will normally be used.

If users are potentially identifiable in publications then they will have explicitly agreed to this (and in some case, e.g. experts or public figures) may have explicitly requested it.

How and **when** (if ever) will this data (or materials) be **reused**

The focus here is on re-use by the original research team. Be realistic.

Note that if you are collecting the data then this will need to be consistent with your information sheet and consent form.

How and **when** (if ever) will this data (or materials) be: **archived, indexed, published** or made available to others?

The focus here is on long-term and discoverable preservation of research data beyond the initial research project.

First, consider the data as it appears in the final output, i.e. thesis or publication:

PhD theses must be submitted to the University's online thesis archive and will be publicly accessible (either immediately or occasionally after an embargo period).

All research publications are effectively publicly available.

Second, consider any data that 'underlies' any published output, e.g. the full transcripts of interviews behind selected quotes, etc.

Increasingly publishers and research funders expect the data 'underlying' a publication to be made available in addition to the publication itself. This may not include all of the collected data (e.g. because of the risk to participants or their non-consent to this, or intellectual property concerns).

Third, consider the complete collected data set (or any significant portions of it) that might be directly useful in someone else's research.

Publicly funded research projects have a moral and contractual obligation to make their data as widely available as possible, bearing in mind the costs and benefits of doing so. But be realistic about whether it will actually be useful to someone else. And again, this may not include all of the collected data (e.g. because of the risk to participants or their non-consent to this, or intellectual property concerns).

See the UoN page on research data management:

https://www.nottingham.ac.uk/fabs/rgs/research-data-management/data-sharing-and-archiving/data-sharing-and-archiving.aspx

Note that the University provides a public research data repository, https://rdmc.nottingham.ac.uk/

How and **when** (if ever) will this data (or materials) be **deleted or destroyed**?

Data relating to PhD student's projects should usually be retained at least until after the degree is awarded. (But see notes on publication, below)

For other research all data should usually be retained for a minimum of 7 years, in case a challenge arises about the validity of the research.[‡]

Any subset of the data that is published may need to be retained for longer.

See the UoN page on research data management:
https://www.nottingham.ac.uk/fabs/rgs/research-data-management/data-sharing-and-archiving/data-sharing-and-archiving.aspx

Note that there is a tension between minimising the use and retention of personal data and retaining data for future use and/or supporting publications or against claims of misconduct. There may be discipline-specific norms for retaining all or only (e.g. anonymised) parts of the data.

After this time, you should plan to delete or destroy the data.

Note that it may be appropriate to retain different parts of the data for different lengths of time, hence to delete or destroy different parts at different points in time.

If human subjects are involved, then at what point(s) can they **withdraw** and what will happen in each case?

In general, you won't be collecting any data from the moment the person withdraws.

In general, it is best practice to allow retrospective withdrawal, i.e. to remove as far as possible data captured or provided prior to their withdrawal.

[‡] https://www.nottingham.ac.uk/governance/records-and-information-management/records-management/retentionschedule.aspx

However, this is not always possible. E.g. after work is published it cannot typically be changed, and data cannot be removed from back-ups. In some cases (e.g. medical trials) it may even be illegal to remove data.

So, the important thing is to be honest with participants. For example, that you will not make any further use of their data after they withdraw but will be unable to remove data from anything already published.

What will happen to this data if/when you **leave the University**?

In general, the requirements about retaining data (above) apply irrespective of whether you are still at the University or not.

If you leave, then who will be responsible? If you are a student, it will usually be your supervisor. If you are a staff member it may be your line manager, the PI of the research project or another relevant staff member.